ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0036

Prediction and Analysis of Stock By PCA-GRU Model Based on Attention Mechanism

Wenjun Zhang

Institute of statistics, Shandong Institute of Business and Technology, Shandong, 264003, China

Abstract

In the stock prediction model established by neural network correlation method, due to the correlation between the selected variables, the data outliers have a great influence on the training results and other factors, resulting in poor generalization and poor prediction results. Based on this problem, this paper makes the closing price prediction for China Merchants Bank's stock, and uses the traditional principal component analysis method to compare the data after the reduction with the relevant technical indicators KDJ, MACD, etc. into different neural network models. Experiments show that the PCA-GRU model based on attention mechanism reduces the prediction error and improves the prediction stability and has practical application value in predicting the closing price trend of China Merchants Bank.

Keywords

Attention mechanism; Neural network method; Principal component analysis method; Stock forecasts.

1. Introduction

With the rapid development of the economy and the maturity of financial markets, financial management behavior is also increasingly popular in the vast number of families, as a development of more than 400 years of non-reimbursable securities - stocks, has also become the majority of investors in the heart of the first choice. At the same time, the trend of stock ups and downs reflects the current economic situation to some extent, so the reasonable analysis of the trend forecast and related laws of stocks is of great significance to both investors and decision makers. However, the high noise, nonlinearity and volatility of stock prices make the accurate prediction of stocks still a difficult problem. In particular, China's stock market is greatly affected by policy, the accurate prediction of stocks is still a great challenge to be explored.

Nowadays, with the continuous development of artificial intelligence technology, the relevant technology is applied in the field of stock prediction, which has greatly improved the accuracy of stock price. At present, more and more scholars use artificial intelligence-related technology to predict stock prices. Whether it is the traditional artificial neural network to deep learning model, or a variety of deep learning neural network combination, artificial intelligence-related technology has been widely used and improved in the stock. Xu Xingjun [1] used the reverse propagation neural network model to predict the stock data of Pudong Development Bank. Applications such as Zhang Chenxi [2] and other support vector machines predict future trend changes in the stock market. Experimental studies, such as Chen's [3], have shown that BP and SVM have different predictive capabilities on different data sets. Selvin [4] et al. compared the three models of RNN, LSTM and CNN, and based on the model's performance in the stock price forecast of NSE listed companies, came to the conclusion that CNN performed better in accuracy. Kim [5] et al. combined the generalized self-regression condition differential model (GRACH)

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0036

with LSTM, and experimented on the KOSPI200 index data, and found that the model was 37.2% lower than the average absolute error of the traditional deep feed-forward neural network, which proved that LSTM has applied value in financial data prediction. As a variant of LSTM, GRU [6, 7] has the advantages of simple structure, not easy to fit, and better performance on small data sets, and is more suitable for USU for financial data than natural language processing (NLP) tasks.

In recent years, as a new computer system- attention mechanism, it has been widely used in the field of neural network-related.In the last century, attention mechanisms emerged with the development of computer vision [8], although it was proposed as early as the last century, but little is said, with the continuous improvement of neural network methods [9, 10], it has flourished in the field of machine translation, and gradually in the stock price forecast.

Pang Chao[11] (2018) and others (2018) introduce attention mechanism into the Enconder-Deconder framework of recursive neural networks to accurately extract the central content of the original text, which is more accurate than the traditional model. Ashish Vaswani [12] and others indicate that attention mechanisms can perform better in more tasks. Chen and Ge [13] have built attention-based LSTM models to predict the movement of Hong Kong's share price. Young [14] et al. (2018) indicates that attention mechanisms help improve the computational efficiency of models, and that attention-based models are currently the most advanced models for multitasking. However, a good stock forecasting model also needs better generalization performance, and in the process of systematic and comprehensive analysis of the problem, many indicators need to be considered, Berradi [15], et al. to predict the 29-day data of a stock on the Casablanca Stock Exchange, which uses PCA to reduce the characteristics from eight to six, experimental results show that the prediction accuracy of the neural network is improved by the data results after dimensionality reduction.

In this paper, the original data is applied to the neural network model based on attention mechanism after the main component method is reduced, and the Atention-GRU, Attention-LSTM model and GRU and LSTM model are compared, and the conclusion that the Atention-PCA-GRU method has a better effect on China Merchants Bank's stock price forecast.

2. Methods to Study

2.1. The Index Data

The data obtained in this article is from python's Tushare financial interface package, which obtains data on nine basic indicators: open, close, high, low, previous closing, ups and downs, rises and losses, volume, and turnover. The data is shown in Table 1. Based on the basic data, KDJ, MACD related technical indicators are obtained.

The KDJ indicator is a sensitive and fast technical analysis indicator that sends a buy and sell signal before a stock price rises or falls. The immature random value RSV of the last day of the period is calculated by the ratio relationship between the high, low, and closing prices of the last period, and then the K, D, and J values are calculated according to the sliding average.

K value is the n-day moving average of RSV, and the K line connected by K value is also called fast line, which changes at a moderate speed among the three curves. D value is the n-day moving average of K value. The middle D line changes at the slowest rate and is called the slow line. J value changes the fastest, as an auxiliary observation of the buy and sell signals sent by K line and D line, J line is called ultra-fast line or confirmation line. The three lines on the same coordinate constitute the KDJ index that can reflect the trend of price fluctuation.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0036

$$\begin{cases} \mathrm{RSV} = (C_n - L_n) / (H_n - L_n) \times 100 \\ K = \dfrac{2}{3} K_p + \dfrac{1}{3} RSV \\ D = \dfrac{2}{3} D_p + \dfrac{1}{3} K \\ J = 3 \times K - 2 \times D \end{cases} \tag{1}$$

In the formula, C_n is the closing price of within n days; L_n is the lowest price within n days, H_n is the highest price within n days; K_p, D_p is the K value and D value of the previous day.

MACD is called the similarity moving average. The convergence and separation of the fast and slow moving averages represent the changes in the market trend, which is also a common technical index of the stock. The moving average EMA of fast and slow speed is generally selected on the 12th and 26th days, and the MACD is finally obtained through calculation of their divergence value DIF and the 9-day moving average DEA of divergence value.

$$\begin{cases} EMA_{(n)} = \dfrac{n-1}{n+1} \times PEMA_{(n)} + \dfrac{2}{n+1} C \\ DIF = EMA_{(12)} - EMA_{(26)} \\ DEA = \dfrac{n-1}{n+1} \times PDEA_n + DIF \\ MACD = (DIF - DEA) \times 2 \end{cases} \tag{2}$$

Where, n is the number of days of moving average; C is the closing price of the day; PEMA and PDEA are EMA and DEA of the previous day.

2.2. The Main Component Method Is Used to Reduce the Dimension of the Data

The main idea of Principal Component Analysis(PCA) is to reduce the dimensionality of multiple highly correlated indicators and transform them into a few comprehensive indicators that can represent the majority of indicator information under the condition of minimizing the loss of data information. The degree to which these comprehensive indicators are representative of the original indicators, measured by contribution rate, is generally considered to be above 85%. In addition, in order to ensure the independence of the principal components and eliminate the correlation between indicators as much as possible, the covariance between the principal components was set to zero.

The general steps for the analysis of the main components are as follows:

(1) Data standardization. Let the original data have n objects and p indicators, and these random variables form a matrix X, where I =1,2... ,n; J = 1, 2,... P. The standard deviation standardization formula is as follows:

$$z_{ij} = \frac{x_{ij} - \mu}{\sigma} \tag{3}$$

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0036

Where μ is the mean value and σ is the standard deviation

(2) Find the correlation matrix. The correlation matrix is shown below

$$R = \frac{z^T z}{n-1} \tag{4}$$

(3) Find the eigenroots and eigenvectors of R. P characteristic roots of R, arranged in order from largest to smallest, is the $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p \ge 0$. Then, the eigenvector $T_1, T_2, \cdots T_p$ corresponding to λ_j is obtained, and the corresponding eigenvector is the eigenvalue λ_j, which is the variance of each principal component.

(4) Divide an eigenvalue by the sum of all eigenvalues to get the variance contribution rate of the eigenvector, $\frac{\sum_{j=1}^{m} \lambda_j}{\sum_{j=1}^{p} \lambda_j}$ is the cumulative contribution rate of the first M principal components.

According to the selection rules of this method, the m principal components selected to meet the requirements contain most of the information of the original data. The eigenvalue of the selected principal component should be greater than 1, and the cumulative contribution rate of the first m principal components should reach 85% (in most cases).

Finally, you can get the main ingredient calculation formula is:

$$Y_i = (T_i^*)' X, i = 1, 2, \cdots, m \tag{5}$$

2.3. Neural Network Model Based on Attention Mechanism

The neural network model used in this paper is long and short-term memory neural network model (LSTM) and gated cyclic neural unit (GRU) are both variants of cyclic neural network (RNN). LSTM is more suitable for dealing with long-distance timing problems. Compared with RNN model, GRU model is a further optimization of LSTM model and more concise and efficient than LSTM model.

Attention mechanism originated from the study of human attention, due to the limitation of information processing ability, people will need to focus on the part of the selective attention information, recursive neural network model to deal with time series data is similar to a markov process, t time step of hidden state h_t and t-1 time only about the hidden state h_{t-1} step, Its conditional distribution function satisfies the following formula:

$$P\{X(t) \le x \mid X(t_n) = x_n, \cdots, X(t_1) = x_1\} = P\{X(t) \le x \mid X(t_n) = x_n\}$$

Although the hidden state h_{t-2} of t-2 time step can affect h_t by influencing h_{t-1}, and so on, the hidden state h_{t-2} of t-n time step can indirectly affect h_t, in fact, such influence has been minimal with the iterative transmission of the network, that is, it is difficult for the neural network to extract the features of the information long ago. The attention mechanism can effectively solve this problem. The attention mechanism will save each hidden state and generate a sequence $(h_1, h_2, \cdots h_t)$, when training, the final output will be associated with this sequence, and selective learning will be carried out according to the correlation with different hidden loading. Figure 1

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0036

shows the operation principle of the attention mechanism, and its expression is shown in Equation (6) :

$$u_t = v^T \tanh(w * h_1 + b_w) \tag{6}$$

Where, v^T and w represent the weights, and u_t represents the average optimal vector corresponding to the current time step t. Therefore, the attention mechanism is essentially A measure of similarity. If the hidden state of certain time step is more similar to the final output value, the greater the weight of the hidden state will be, and the more contributions it will make to the model. In other words, the attention mechanism effectively solves the limitation of time length on the model, and the hidden state of each time step has equal opportunity to participate in the final model interpretation. Meanwhile, the output is directly associated with all the reference timeseries data to obtain more accurate output results.

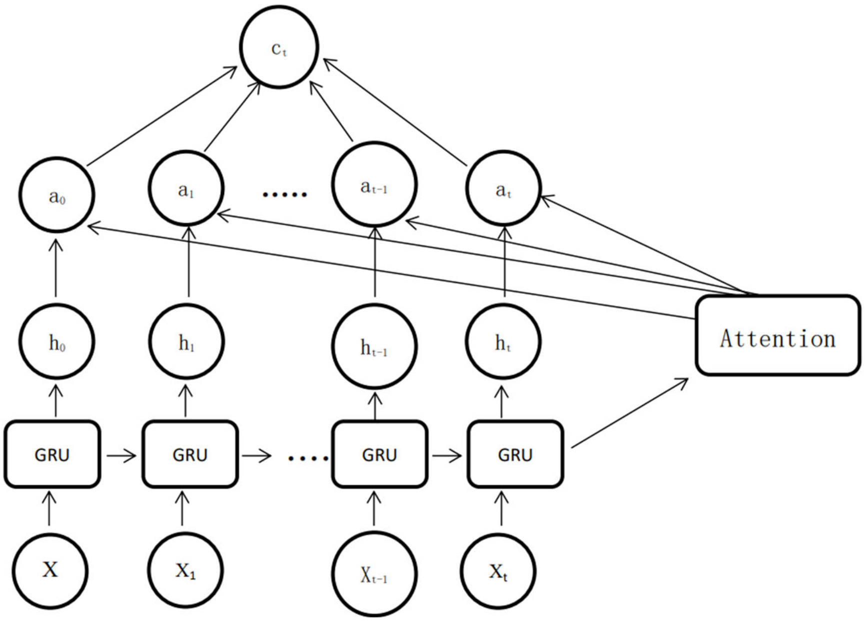

Figure 1. Schematic diagram of GRU model based on Attention Mechanism

Similarly, the Attention mechanism can be added to the LSTM based model, and the Attention-based LSTM and Attention-based GRU neural network models are constructed respectively. In the empirical analysis in Chapter 4, by comparing with other models, The effect of attention mechanism on model performance is discussed in detail.

3. The Example Analysis

3.1. Sample and Data Dimensionality Reduction

Based on the principle of representativeness, China Merchants Bank, one of the three giants in The Shanghai Composite stock index, was selected with stock code 600036.SH. 4,640 groups of sample data from April 10, 2002 to August 31, 2021 were selected. In terms of experimental Settings, 80% of the data were divided into training sets and 20% of the data were divided into test sets. Using python's Tushare financial data interface package, we get nine basic data about the selected stock: open price, closing price, high price, low price, previous closing price, up/down amount, turnover rate, volume, and transaction amount.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0036

In the model based on attention mechanism, the raw data are further processed to calculate the random index values of cycle RSV, DIF, DEA, etc., and the obtained daily KDJ, MACD and other related indicators are used as training data.

According to the principal component analysis method and formula (1), the original data were standardized to obtain the correlation coefficient, and the characteristic equation was solved to obtain the eigenvalue and variance contribution rate, as shown in Table 1:

As can be seen from Table 1, the cumulative variance contribution rate of the first three components is 98%, and the characteristic value is greater than 1, so the first three main components are selected. By further calculation, the component load values of the first three components, such as Table 2, are obtained.

Based on the component load value of Table 2 and the variance contribution rate of the three main components of Table 1, the coefficient of the three main components is calculated, and the coefficient result is shown in Table 3, so as to obtain the formula for calculating the following three main components as shown in formula (7).

Table 1. Eigenvalue and variance contribution rate

composition	Initial eigenvalue			Extract the sum of squares and load		
	Total	Percentage of variance	The cumulative %	Total	Percentage of variance	The cumulative %
1	5.145	57.165	57.165	5.145	57.165	57.165
2	2.107	23.411	80.576	2.107	23.411	80.576
3	1.609	17.882	98.459	1.609	17.882	98.459
4	0.111	1.236	99.695			
5	0.026	0.293	99.988			
6	0.001	0.007	99.95			
7	0.000	0.004	99.998			
8	0.000	0.002	100.000			
9	-1.006E	-1.062E-013	100.00			

Table 2. Component load matrix

	component 1	component 2	component 3
1	0.995	-0.069	0.064
2	0.995	-0.061	0.082
3	0.994	-0.091	0.058
4	0.994	-0.099	0.051
5	0.994	-0.082	0.075
6	0.022	0.782	0.576
7	0.031	0.758	0.609
8	0.124	0.699	-0.695
9	0.430	0.632	-0.633

Table 3. Correlation coefficients of principal components

	factor 1	factor 2	factor 3
1	0.439	-0.048	0.050
2	0.439	-0.042	0.065
3	0.438	-0.063	0.046
4	0.438	-0.068	0.040
5	0.438	-0.056	0.059
6	0.100	0.539	0.454
7	0.014	0.522	0.480
8	0.055	0.482	-0.548
9	0.190	0.435	-0.500

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0036

$$
\begin{aligned}
Y_1 &= 0.439X_1 + 0.439X_2 + \cdots + 0.190X_9 \\
Y_2 &= -0.048X_1 - 0.042X_2 + \cdots + 0.435X_9 \\
Y_3 &= 0.050X_1 + 0.065X_2 + \cdots - 0.5X_9
\end{aligned} \tag{7}
$$

The three main component results and KDJ, MACD and other technical indicators, together as a new data indicators to establish Atention-PCA-GRU model and other network models.

3.2. Experimental Settings and Metrics

Based on the selection of super parameter and algorithm optimization in neural network method, the selection of parameters in neural network model and the setting of learning algorithm in training network will be introduced. In the process of updating and optimizing the internal parameters of the model using training data, the random gradient drop method (SGD) is selected as the iterative learning algorithm. The algorithm makes the loss function value of the optimization parameter as small as possible, and uses the reverse propagation update algorithm to realize the algorithm optimization. The reverse propagation algorithm is used. The most important super parameters are The Batch size and the number of epochs are in a series of processes in which the loss function is calculated by algorithmic prediction and real results. Batch size indicates that the training sample is divided into batches for training tuning. The number of Batches indicates how many samples are trained for a batch of models, and different models with different Batch sizes get different predictions. When batch size is a sample, the learning algorithm is the random gradient drop algorithm. When the batch size exceeds one sample and is smaller than the size of the training dataset, the learning algorithm drops for small batch gradients. One or more Batches form an Epoch. The number of Epochs represents the number of times the entire training data or number of times it has been traversed, as well as the number of times all batches have been trained and internal parameters updated. In this paper, a small batch gradient drop algorithm is selected to train the model.

In this paper, the data for the first five days of the training set is entered as features, and the closing price of the sixth day is labeled and trained by the model. The number of model variables under the original data is 16, so the input layer dimension is 16, according to the main component analysis of the data dimension is 10, then the main component analysis after the input layer dimension is 10. The output layer dimension is 1. Because neural networks continuously improve prediction accuracy based on the error reverse conduction of prediction values and real values, several different evaluation criteria are considered. This paper uses SEVERAL, RMSE and MAE metrics. MAPE is an average absolute percentage error, a percentage error, and its numerical results can be greater than 100 percent. Root Mean Square Error is a measure of the deviation between the predicted and real values, which can make the experimental results better described and MSE under the root number. Mean Absolute Error is also known as L1 fan loss. Although MAE can better measure the quality of regression model, but the existence of absolute values lead to the function is not smooth, at some points can not be guided. The measure is not 10 1000, can not just say that a measure on the effect of good, that the model performance is good, so here we use these three measures to compare the effect of the model.

Suppose y_t is the true value at time t, $\hat{y}_t$ is the predicted value at time t, MAPE is defined by $MAPE = \frac{100\%}{n}\sum_{i=1}^{n}|\frac{\hat{y}_i - y_i}{y_i}|$,MSE is defined by $MSE = \frac{1}{N}\sum_{i=1}^{N}(y_t^i - \hat{y}_t^i)^2$ RMSE is the square root of MSE, and MAE is expressed by $MAE = \frac{1}{N}\sum_{i=1}^{N}|y_t^i - \hat{y}_t^i|$.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0036

3.3. Comparative Analysis of Experimental Results

In order to reflect the effectiveness of the Attention-based PCA-GRU model in China Merchants Bank's share price forecast. Compare the prediction effect and error of the LSTM, GRU, Attention-LSTM, Attention-GRU models under the original data. Because the neural network model based on attention mechanism fits significantly well, the Atention-LSTM and Atention-GRU models are further selected to train the data after the main component is de-dimensional. The true and predicted values of the LSTM network under the original data are line charted in Figure 2, and the prediction results of the GRU, Attention-LSTM, Attention-GRU models are Figure 3, Figure 4, and Figure 5, respectively. The forecast error is shown in Table 4.

From Figure 2 to Figure 4 forecast fit trend chart can be seen, China Merchants Bank stock data indicators, based on the attention mechanism of the neural network of the stock closing price forecast trend situation fits better, the figure based on the attention mechanism network forecast value and the real value trend is basically the same, and significantly better than the LSTM network and GRU network, the two networks forecast trend line and the real trend line difference is larger, the degree of fit is poor. At the same time, as can be seen from Table 4, the measurement of LSTM network and GRU network is quite different from the neural network error index based on attention mechanism. The MAP, MAE and RMSE values of neural networks based on attention mechanism are significantly smaller than the errors under neural networks, which further illustrates the validity of the neural network model based on attention mechanism on the trend forecast of China Merchants Bank's stock price. On the other hand, according to the comparison between the attention-based neural networks based on the mechanism of attention in Figure 3 and Figure 4, the Atention-GRU model is better fitted than the Atention-LSTM model in the original data. At the same time, the prediction error of the Atention-GRU model in Table 4 is smaller than that of the Atention-LSTM network, which indicates that the Atention-GRU model is more accurate for the forecast of china Merchants Bank's stock price trend, and in order to further improve the effect of the stock price forecast, the neural network model based on the attention mechanism will be used to predict the reduction data under the main component, and the prediction effect and error will be compared.

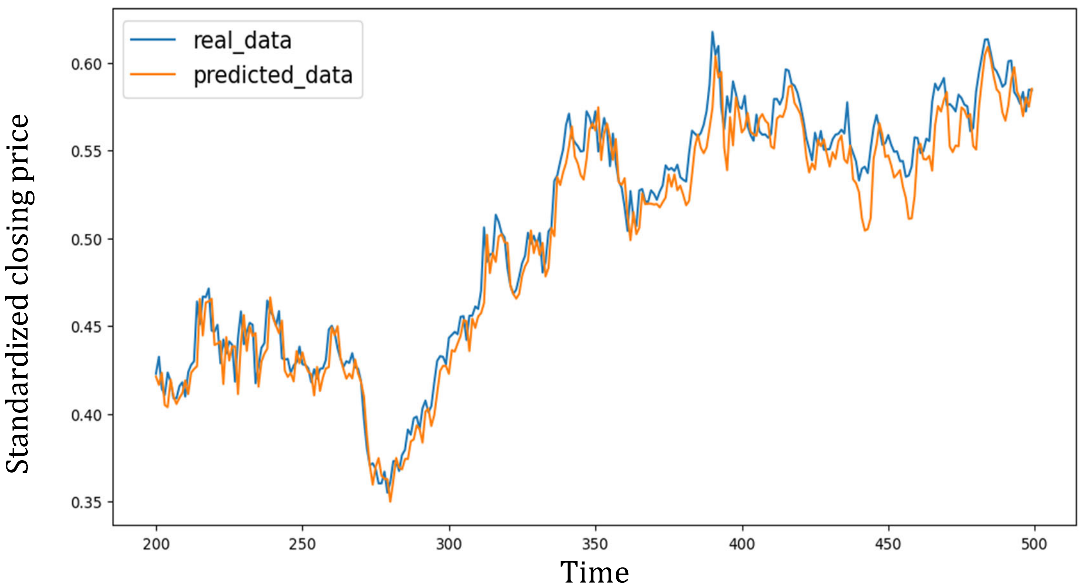

Figure 2. LSTM model fitting results

ISSN: 2637-6067
DOI: 10.6918/IJOSSER.202112_4(12).0036

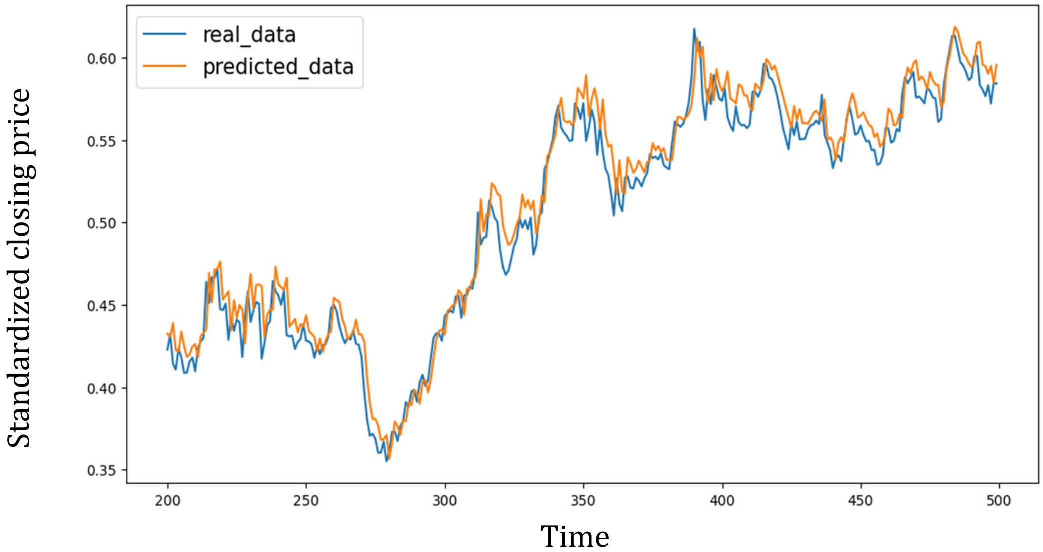

Figure 3. GRU model fitting results

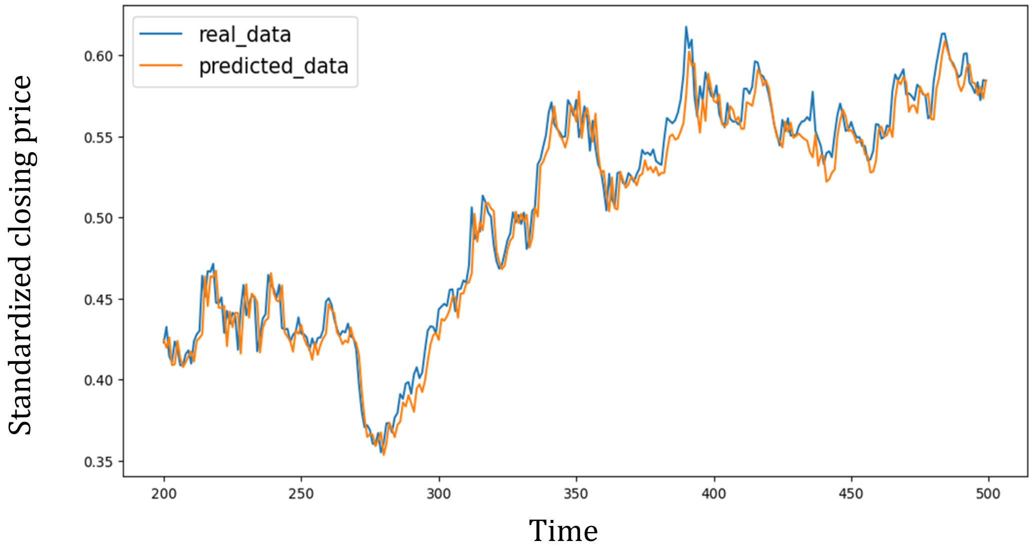

Figure 4. Fitting results of Attention+LSTM model

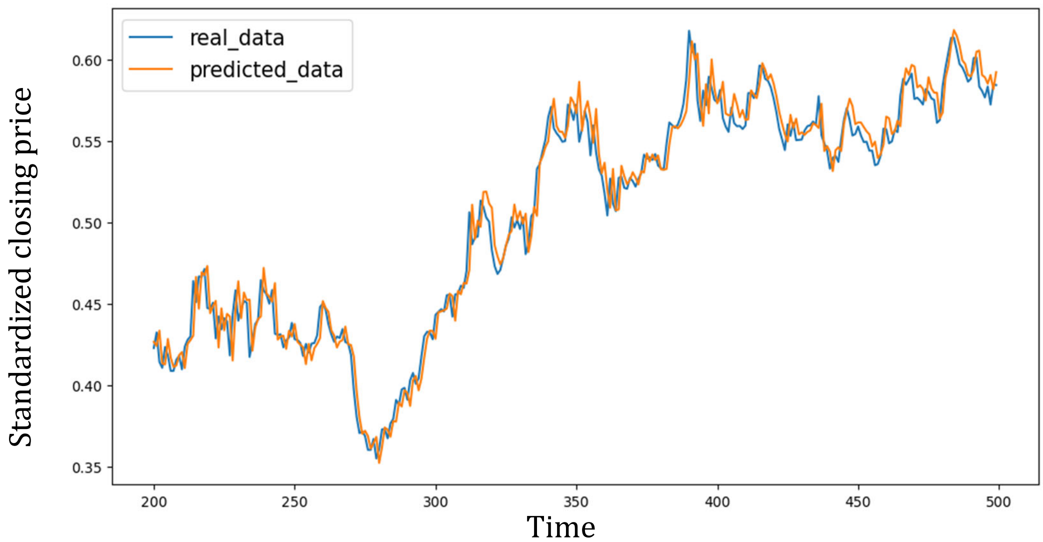

Figure 5. Prediction results of Attention+GRU model

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0036

Table 4. Prediction error results of each model

The model	The evaluation index		
	MAPE	MAE	RMSE
GRU	4.7932	0.0301	0.0473
GRU	4.1689	0.0270	0.0428
Attention+LSTM	3.7646	0.0248	0.0404
Attention+GRU	2.7002	0.0178	0.0294

Figure 5 and Figure 6 are line charts of the two neural network models based on attention mechanism after the main component is reduced, and Table 5 is the fitting error between the original data and the main component after the reduction of dimension in the two models, the results are shown below.

According to the analysis results of different data dimensions in Table 5, after the main component is reduced, the MAP, MAE and RMSE values of the attention-based LSTM model and the GRU model are smaller than the error values under the original data, indicating the validity of the main component method to the data reduction of the relevant index data of China Merchants Bank, and the effect of the main component method on the prediction of stock price trend after combining with the attention-based neural network. Similar to the prediction effect under the original data, the prediction curve of the GRU network model based on the Attention mechanism coincides better with the real curve, and the prediction error of the data after the descent is still smaller than that of the LSTM network model based on the attention mechanism. Thus, it shows the remarkable validity of the PCA_GRU composite model based on attention mechanism to China Merchants Bank's share price forecast.

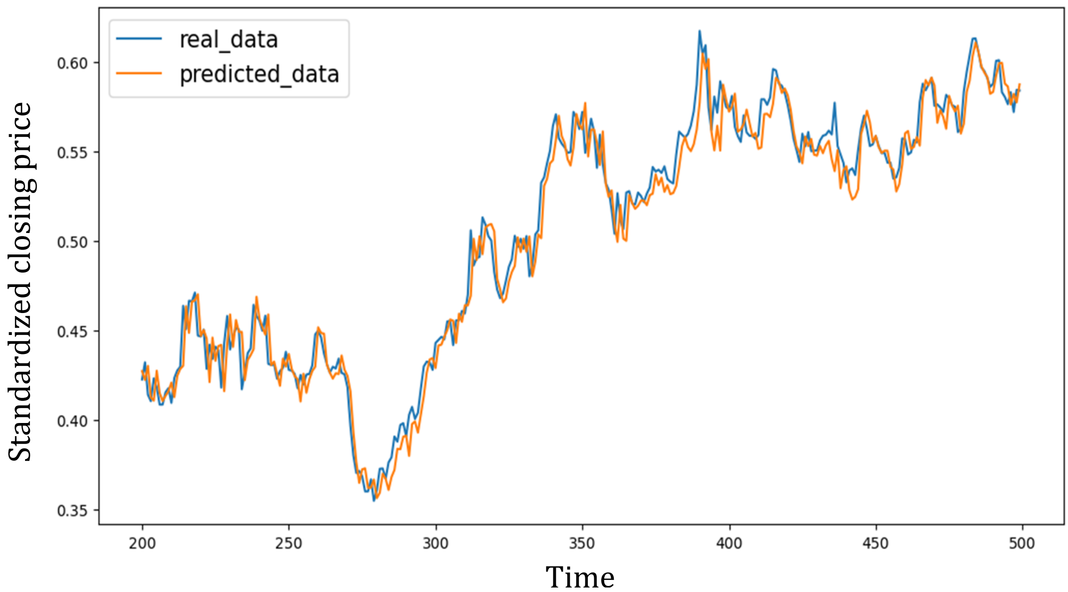

Figure 6. Fitting results of PCA_LSTM model based on attention mechanism

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0036

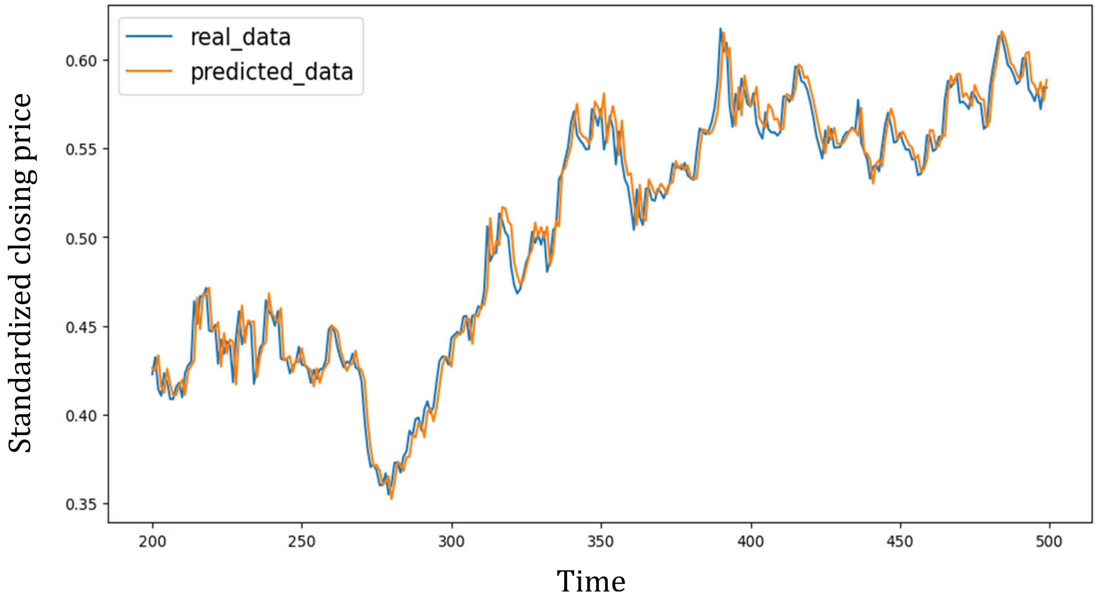

Figure 7. Fitting results of PCA_GRU model based on Attention Mechanism

Table 5. Model prediction error results under principal component method

The model	The evaluation index		
	MAPE	MAE	RMSE
Attention+LSTM	3.7646	0.0248	0.0404
Attention+GRU	2.7002	0.0178	0.0294
Attention+PCA_LSTM	3.1523	0.0201	0.0362
Attention+PCA_GRU	2.4782	0.0162	0.0263

4. Conclusion and Prospect

This paper mainly uses the comparison of various model methods to extract the three main components of the original data by using the main component analysis method, and adds the stock-related technical indicators to improve the sample quality. The correlation between the co-relationship features improves the modeling effect of neural network methods on data and the specific stock data by digging deep into the information implied in the data. Through the control experiments between various neural networks and the network based on attention mechanism, and then the simulation fit of the model after the analysis of the main components, the PCA-GRU model based on the attention mechanism is obtained which has a significant effect on improving the accuracy of China Merchants Bank's closing price prediction.

Although the results are biased from the true values, the overall trend predicted is consistent. However, the stock market is susceptible to external factors and its own instability, although the research of this paper has preliminary conclusions, but there are still some problems that need to be improved and noticed:

(1) Many models in the neural network are specific to different data conclusions, different algorithms under different data and super parameter selection and so on. In this paper, only one stock is selected for comparative analysis on different models, and the fitting effect of models on other stocks or data sets should be further considered.

(2) The stock markets have always been more complex and volatile, in addition to the relevant indicator data selection has an impact on the forecast results, relevant national policies,

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0036

industry development and human intervention are all factors affecting the trend of stocks. Therefore, more and more scholars use text analysis to predict stocks.

(3) The measure price index reflects the stock market lag, the model can not predict the emergency in time. At the same time, most of the stock forecast is based on historical data up and down the trend of the fit, for accurate prediction of stock prices and so on is still a huge challenge, based on this, stock research has a long way to go.

References

[1] Xu Xingjun, Yan Gangfeng. Stock Price Trend Analysis Based on BP Neural Networks, Zhejiang Finance, 20119 (11): 57-59-64.

[2] Zhang Chenxi, Zhang Yanping, Zhang Yingchun, Chen Jie, Wan Zhong. Based on the stock forecast of support vector machines. Computer Technology and Development, 2006 (06): 35-37.

[3] Chen W H,ct al.Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets[J].International Journal of Electronic Finance,2006,1(1):4~67.

[4] Selvin S, Vinayakumar R, Gopalakrishnan E A, Stock price prediction using LSTM, RNN, and CNN-SLIDING window model [C] //Proc of International Conference on Advances in Computing, Communications and Informatics. Piscataway, NJ:IEEE Press, 2017: 1643-1647.

[5] Kim H Y, Won C H. Forecasting the volatility of stock price index: a hybrid model integrating LSTM with multiple GARCH-type models [J]. Expert Systems with Applications, 2018, 103(8):25-37.

[6] Chung J Y, Gulcehre C, Cho K H. Empirical evaluation of gated recurrent neural networks on sequence modeling [C]//Proc of Deep Learning and Representation Learning Workshop.2014: 1026-1032.

[7] Cho K,Van Merrienboer B,Gulcehre C,et al.L-earning phrase representations using RNN encoder-decoder for statistical machine translation [J]. Computer Science, 2014, 8(3):15-23.

[8] Mnih,V.;Heess,N.;Graves,A.;Kavukcuoglu,K.Recurrent Models of Visual Attention.Advances in Neural Information Processing Systems.arXiv 2014,arXiv:1406.6247.

[9] Bahdanau,D.;Cho,K.;Bengio,Y.Neural Machine Translation by Jointly Learning to Align and Translate.arXiv 2015,arXiv:1409.0473.

[10] Vaswani,A.;Shazeer,N.;Parmar,N.;Uszkoreit,J.;Jones,L.;Gomez,A.N.;Kaiser,L.;Polosukhin,I.Attention is All you Need.arXiv 2017,arXiv:1706.03762.

[11] Chao, Yu Chuanhuan. Classification-based Chinese Text Summary Methods .J.Computer Science, 45(1): 144-147, 2018.

[12] Ashish Vaswani.Noam Shazeer.Niki Parmar.Jakob Uszkoreit.Llion Jones.Aidan N.Gomez.Lukasz Kaiser.Attention Is All You Need.3 lst Conference on Neural Information processing System(NIPS 2017),Long Beach,CA,USA.

[13] Chen, S.; Ge, L. Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction. Quant. Financ.2019, 19, 1507–1515. [CrossRef]

[14] Tom Yong,Devamanyu Hazarika,Soujanya Poria,and Erik Cambria.Recent trends in deep learning based natural language processing.IEEE Computational Intelligence magazine, 13(3):55-75,2018.

[15] Berradi Z,Lazaar M.Integration of principal component analysis and recurrent neural network to forecast the stock price of Casablanca stock exchange [J]. Procedia Computer Science, 2019, 148:55-61.